Minimization Algorithms for the Maximal Entropy Method

Antal Jakovac

University of Wuppertal
BME Technical University Budapest

April 23 2008.

Outlines

Outlines

Spectral functions:

- spectrum – density of states:
  - peaks: quasiparticles in a given quantum channel
  - peak width: decay constant or inverse lifetime
  - peak height: form factors, thermal abundance
  - continuum: collective phenomena
- linear response theory: imaginary part of Green functions, by Kramers-Kronig relation complete retarded Greens function can be reconstructed – real time evolution, transport coefficients

most reliable calculation of the spectral function?

Nonperturbative method: MC lattice simulations

- discrete Euclidean spacetime $\Rightarrow$ provides Euclidean propagators
- to obtain spectral function we can use exact relations but it needs analytic continuation – impossible from lattice data
- statistical analysis is ill-defined, because the number of parameters (now value of the spectral function at each frequency) $\gg$ number of data
- solution: build in prior knowledge into the statistical analysis

best tool: Maximal Entropy Method (MEM)

- relatively new in statistical analysis in high energy physics

  Y. Nakahara, M. Asakawa and T. Hatsuda, Phys. Rev. D 60, 091503 (1999)
  M. Asakawa, Y. Nakahara and T. Hatsuda, Progress in Particle and Nuclear Physics 46 (2001) 459-508.
  F. Karsch, E. Laermann, P. Petreczky, S. Stickan and I. Wetzorke, Phys. Lett. B 530, 147 (2002)
  M. Asakawa and T. Hatsuda, Phys. Rev. Lett. 92, 012001 (2004)
  S. Datta, F. Karsch, P. Petreczky and I. Wetzorke, Phys.Rev. D 69, 094507 (2004)
  T. Umeda, K. Nomura and H. Matsufuru, Eur.Phys.J. C39S1 (2005) 9-26

  G. Aarts, C. Allton, M. B. Oktay, M. Peardon, J.-I. Skullerud, Phys.Rev. D 76, 094513 (2007)

- algorithmic improvement, test, and obtain heavy quarkonium spectral functions at zero and finite temperature

  A.J., P. Petreczky, K. Petrov, A. Velytsky, Phys.Rev. D75, 014506 (2007)

Outlines

Nonperturbative method: MC lattice simulations

- Euclidean propagators
  $D_A(\tau, \mathbf{x}) = \langle \mathcal{T}_\tau A(\tau, \mathbf{x}) A(0) \rangle$, where $A^\dagger = A$ self-adjoint bosonic operator, $\tau = j \times a$, $j = 1, 2, \ldots N_t$.
- projection on representation of the spatial rotation group
  $D_{A,J}(\tau) = \sum_{\mathbf{x}} R_J(\mathbf{x})\, D_A(\tau, \mathbf{x})$.

How do we read off the spectral function?

Definition of the Euclidean spectral function & KMS relation:

$$D_A(\tau, \mathbf{x}) = \int_0^\infty d\omega\, K(\tau, \omega)\sigma_A(\omega, \mathbf{x}), \qquad \text{where}$$

$$K(\tau, \omega) = \frac{e^{(\beta - |\tau|)\omega} + e^{|\tau|\omega}}{e^{\beta\omega} - 1} = \frac{\cosh(\frac{\beta}{2} - |\tau|)\omega}{\sinh \frac{\beta}{2}\omega}$$

$\Rightarrow$ exact relation!

After projection

$$D_{A,J}(\tau) = \int_0^\infty d\omega\, K(\tau, \omega)\sigma_{H,J}(\omega).$$

Goal: invert this relation

For $\tau > 0$ it is a (bilateral) Laplace transform:

$$D_A(\tau > 0) = \int_{-\infty}^{\infty} d\omega \, e^{\tau\omega}(1 + n(\omega))\sigma_A(\omega),$$

If we could analytically extend it to complex arguments, then the inverse transformation would go with the Bromwich-integral:

$$(1 + n(\omega))\sigma_A(\omega) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} dt \, e^{-\omega t} D_A(t).$$

Analytic continuation from a finite number of points is impossible.

Outlines

Instead: we treat the measured values of D_A as constraints, and look for the best σ that reproduces these values

$\Rightarrow$ statistical, χ^2 method.

Likelihood of having σ_A spectral function $\mathcal{P}(\sigma_A) \sim e^{-\frac{1}{2}\chi^2[\sigma_A]}$ where

$$\chi^2[\sigma_A] = \sum_{\tau,\tau'} [D_A(\tau) - \bar{D}_A(\tau)]\, C^{-1}_{\tau\tau'}\, [D_A(\tau') - \bar{D}_A(\tau')],$$

where

- $\bar{D}_A$ are the measured propagator values
- $D_A(\tau) = \int K(\tau,\omega)\sigma_A(\omega)$, the calculated propagator values,
- $C_{\tau,\tau'}$ is the correlation matrix of the measurements.

Problem: we know D_A at most at $\mathcal{O}(10-100)$ points, and we need σ_A at $\mathcal{O}(1000)$ points $\Rightarrow$ χ^2 method is ill-defined.

Solutions

- Ansatz for the spectral function – hard to find a reliable one!
- choose a functional basis

  H.B. Meyer, Phys.Rev. D76, 101701 (2007)
- weight the values of σ_A with some pre-defined distribution $p(\sigma_A)$ (prior probability)

$$\Rightarrow \quad \mathcal{P}(\sigma_A) \sim e^{-\frac{1}{2}\chi^2[\sigma_A]} p(\sigma_A)$$

Maximal Entropy Method (MEM) is based on this approach.
More mathematical description is based on Bayesian analysis

M. Asakawa, Y. Nakahara and T. Hatsuda, Progress in Particle and Nuclear Physics 46 (2001) 459-508.

Weight probability is especially appropriate tool to take into account prior knowledges. Now

$$\sigma_A \quad \text{is real and} \quad \sigma_A(\omega > 0) > 0.$$

$\Rightarrow$ weight probability should allow only positive values.
Candidate: (continuous) Poisson distribution with pre-defined positive averages:

$$\frac{\kappa^n}{n!} e^{-\kappa} \to \left\{ \begin{array}{l} n \to \alpha\sigma_A(\omega) \\ dn \to \alpha d\sigma_A(\omega) \\ \kappa \to \alpha m(\omega) \end{array} \right\} \to \sqrt{\frac{\alpha}{2\pi\sigma_A}} e^{-\alpha S_{SJ}[\sigma_A, h]},$$

$$\text{where} \qquad S_{SJ}[\sigma_A, h] = \int d\omega \left[\sigma_A(\omega) \ln \frac{\sigma_A(\omega)}{m(\omega)} - \sigma_A(\omega) + m(\omega) \right],$$

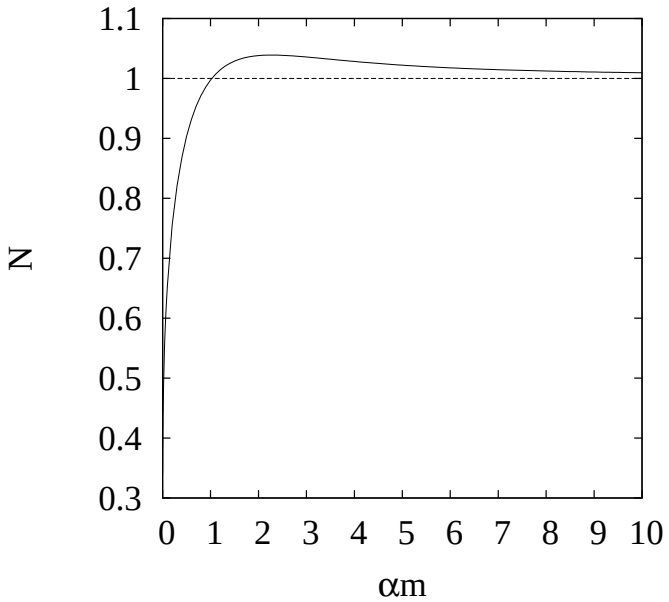
also known as the Shannon-Jaynes entropy.
$m(\omega)$ is called default model.

Integration measure: $[d\sigma_A] = \prod_\omega d\sigma_A(\omega) \sqrt{\frac{\alpha}{2\pi\sigma_A(\omega)}}$

To normalize

$$\frac{1}{\prod_\omega \mathcal{N}(\alpha m(\omega))} \int [d\sigma_A] e^{-\alpha S_{SJ}[\sigma_A, h]} = 1,$$

where



New likelihood function $Q = \frac{1}{2}\chi^2 + \alpha S_{SJ}$

- α controls the relative importance of the data and the weighting distribution
- $\alpha = 0$ pure χ^2 method
- minimum is unique for $\alpha > 0$

Outlines

Discretizing:

$$\omega = \{0 \ldots N_\omega - 1\}\Delta\omega, \quad N_\omega \sim \mathcal{O}(1000)$$
$$\tau = \{0 \ldots N_\tau/2 - 1\}\beta/N_\tau$$

Minimum condition:

$$\frac{\partial Q}{\partial \sigma_A(\omega)} = \alpha \ln \frac{\sigma_A(\omega)}{m(\omega)} + \sum_{\tau,\tau'} K(\tau,\omega) C^{-1}_{\tau\tau'} (D_A(\tau') - \bar{D}_A(\tau')) = 0.$$

where $D_A(\tau') = \Delta\omega \sum_{\omega'} K(\tau',\omega')\sigma_A(\omega')$

- N_ω coupled nonlinear equations
- solution: recursive approximation with local linearization and finding approximate minimum
- number of parameters $\sim$ number of independent vectors $\sim$ rank of the second derivative matrix: number-of-data $(N_\tau/2)$

  $\Rightarrow$ inversion problematic

- Bryan algorithm (R. K. Bryan, Eur. Biophys J. 18, 165 (1990)) uses singular value decomposition to invert the matrix
- number of independent singular vectors$= N_\tau/2$ (G. Aarts, C. Allton, M. B. Oktay, M. Peardon, J.-I. Skullerud, Phys.Rev. D 76, 094513 (2007))
- Our improvement (A.J., P. Petreczky, K. Petrov, A. Velytsky, Phys.Rev. D75, 014506 (2007)):
  - identify the singular vectors exactly
  - reducing the problem to number-of-data dimensional minimization problem

Improvement

- $K(\tau,\omega)$ linear independent for different τs ($\sim \cosh(\frac{\beta}{2}-\tau)\omega$)
- make a basis on N_ω dim. space: $\{R_i(\omega)\}_{i=0,\ldots N_\omega-1}$, where $R_i(\omega) = K(\tau_i,\omega)$ for $i = 0,\ldots \frac{N_\tau}{2}-1$.
- expand $\ln \sigma_A(\omega)/m(\omega)$ in this basis:

$$\ln \frac{\sigma_A(\omega)}{m(\omega)} = \sum_\tau s(\tau)K(\tau,\omega) + \sum_{i\geq N_\tau/2} \tilde{s}_i R_i(\omega).$$

- then we have

$$\sum_{\tau,\tau'} K(\tau,\omega)\left[\alpha s(\tau)\delta_{\tau\tau'} + C^{-1}_{\tau\tau'}(D_A(\tau') - \bar{D}_A(\tau'))\right] + \sum_{i\geq N_\tau/2} \tilde{s}_i R_i(\omega) = 0.$$

- linear independence requires $\tilde{s}_i = 0$
- first $N_\tau/2$ equation yields

$$\alpha \sum_{\tau'} C_{\tau\tau'} s(\tau') + \Delta\omega \sum_{\omega} K(\tau,\omega)\sigma_A(\omega) - \bar{D}_A(\tau) = 0,$$

where $\sigma_A(\omega) = m(\omega)\exp\{\sum K(\tau,\omega)s(\tau).\}$

- This equation can be written as $\frac{\partial U}{\partial s} = 0$ with

$$U[s] = \frac{\alpha}{2}\sum_{\tau\tau'} s(\tau) C_{\tau\tau'} s(\tau') + \int d\omega \sigma_A(\omega) - \sum_{\tau} s(\tau)\bar{D}_A(\tau).$$

$\Rightarrow$ potential minimization problem in $N_\tau/2$ dimensions

Outlines

technical trick: at small ω $\quad K(\tau,\omega) \to \frac{T}{\omega}$. To avoid divergent behavior, rescale by $f(\omega) \sim \omega$ for small frequencies:

$$\tilde{K}(\tau,\omega) = f(\omega) K(\tau,\omega)$$

We obtain $\tilde{\sigma}_A(\omega)$ as a solution.
We reproduce the original spectral function from the condition that it produces the same propagator:

$$\tilde{K}(\tau,\omega)\tilde{\sigma}_A(\omega) = K(\tau,\omega)\sigma_A(\omega) \quad \Rightarrow \quad \sigma_A(\omega) = f(\omega)\tilde{\sigma}_A(\omega)$$

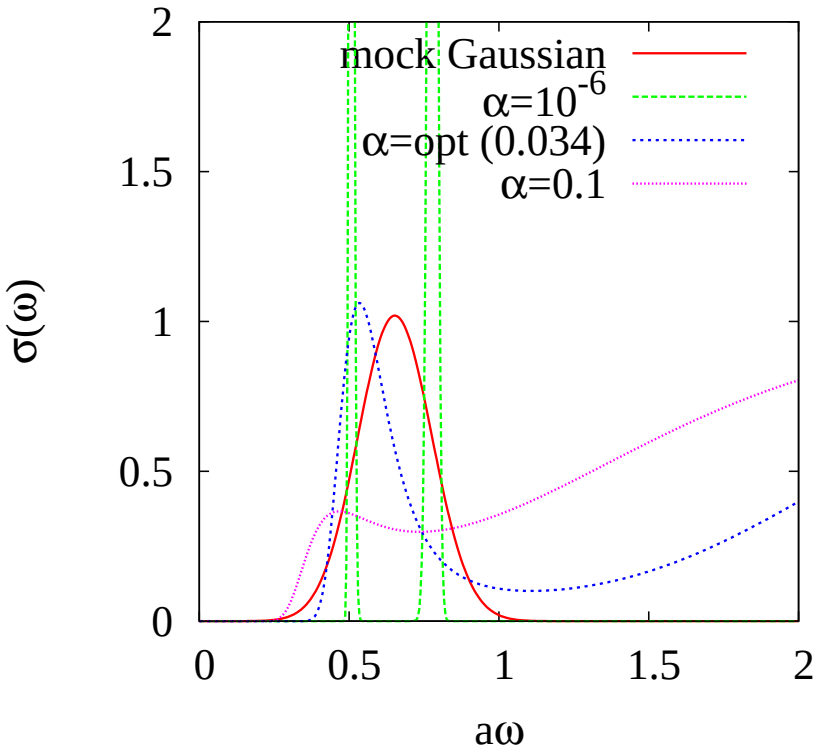
- not too sensitive to the exact choice of $f(\omega)$
- we used $f(\omega) = \tanh(\beta\omega)$

Outlines

How should we treat α in $\mathcal{P}(\sigma_A) \sim e^{-\frac{1}{2}\chi^2 - \alpha S_{SJ}}$?

- at $\alpha = 0$ the problem is ill-defined; numerically $\exists\ \alpha_{\min}$, in our cases typically $\mathcal{O}(10^{-13} - 10^{-20})$.
- $\alpha \neq 0$ introduces a systematic bias into the equation: not the $D_A(\tau) = \bar{D}_A(\tau)$ choice is preferred. This disfavors to large α.
- if α is too small, the spectrum becomes too peaky
- there exists an optimal choice – how do we find it?

Behavior as a function of α: noisy Gaussian curve ($N_\tau = 32$)



- too small α: sharp peaks
- optimal α roughly reproduces the original peak; considerable fake high frequency part
- too large α: too broad distribution

Treat α as a parameter to fit:

- use maximum likelihood to fix σ_A and α
- modified likelihood function:

$$\mathcal{P}(\sigma_A, \alpha) = \mathcal{P}(\alpha) \prod_\omega \sqrt{\frac{\alpha}{\sigma_A}} \exp\left(-\frac{1}{2}\chi^2 - \alpha S_{SJ}\right).$$

  $\mathcal{P}(\alpha)$ may contain an arbitrary α dependence (usual choices are const., or $1/\alpha$) and normalization.

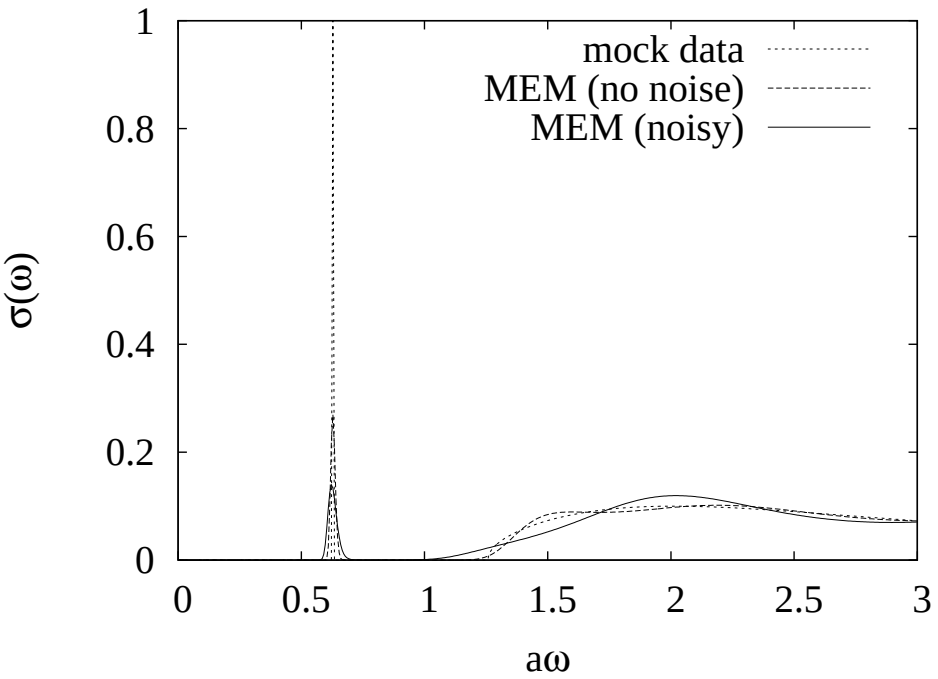
- Usually strongly peaked in σ_A and flatter in α; in saddle point approximation (minimum at $\sigma_A^\alpha(\omega)$):

$$\mathcal{P}(\sigma_A, \alpha) \approx \mathcal{P}(\alpha)\, \delta(\sigma_A - \sigma_A^\alpha)\, e^{\frac{1}{2}\sum_\omega \ln \frac{\alpha}{\lambda_\omega + \alpha} - \frac{1}{2}\chi^2(\sigma_A^\alpha) - \alpha S_{SJ}(\sigma_A^\alpha)},$$

  where λ_ω are eigenvalues of
  $\frac{1}{2}\sqrt{\sigma_A(\omega)}\frac{\partial \chi^2}{\partial \sigma_A(\omega)\partial \sigma_A(\omega')}\sqrt{\sigma_A(\omega')}$ matrix.

Outlines

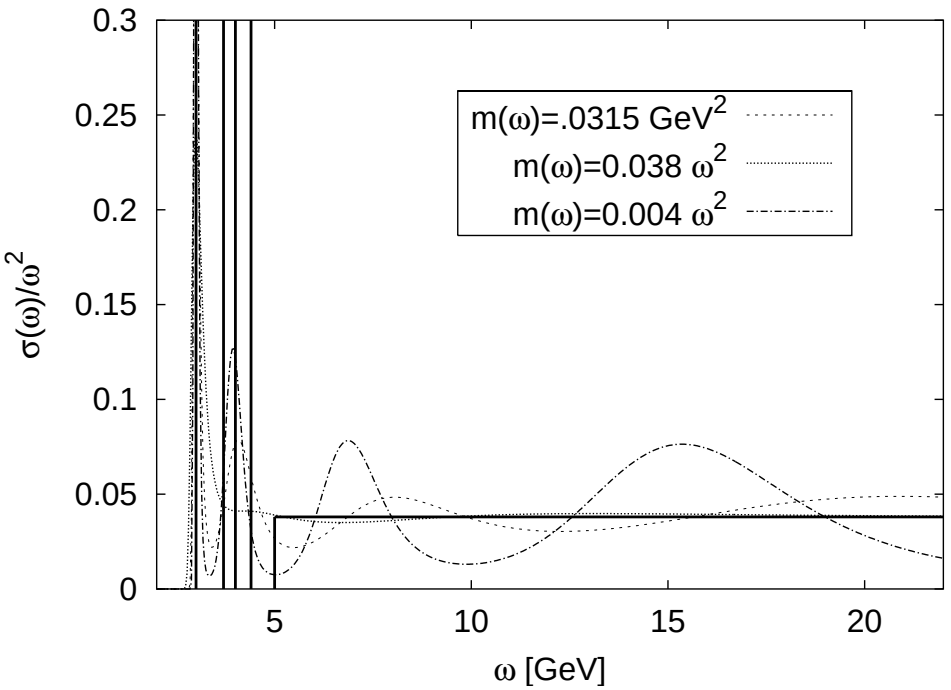
There is statistical error in measurements
Its effect in a typical continuum spectral function ($N_\tau = 32$):



- lowest energy state is well reproduced
- reducing the noise improves the reproduction of the continuum part
- oscillation/peak even when the mock model is flat

A more realistic example with semi-realistic charmonium masses

from A. Mocsy and P. Petreczky, Phys. Rev. D 73, 074007 (2006)

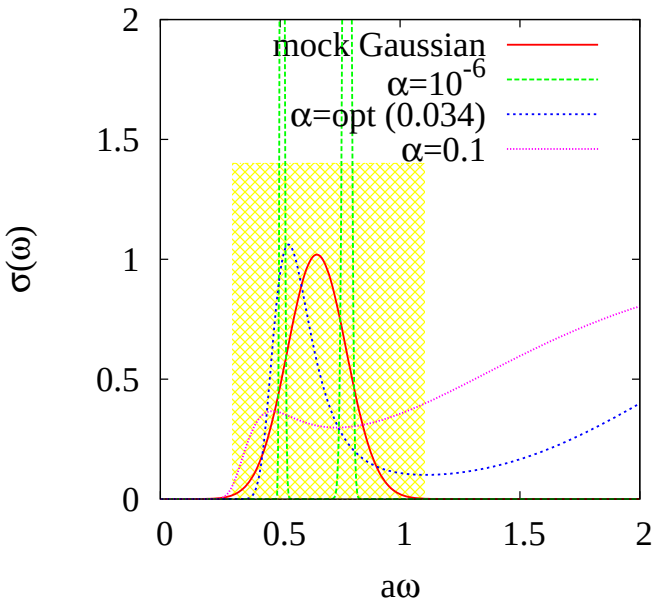


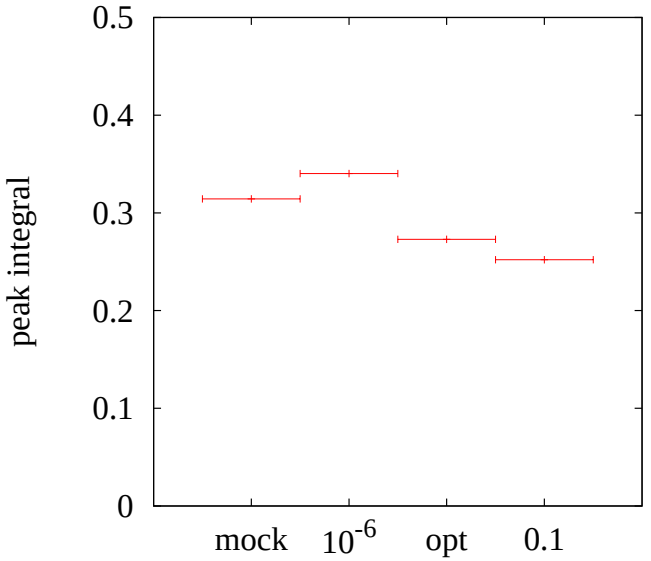
- first peak correct
- three resonances appear as a broad peak
- continuum still contains oscillations

Outlines

Error estimation:

- we want to see the particle peaks
- peak integral is pretty much insensitive to α.
  Example: Gaussian mock data, peak region $\omega a = [0.3 : 1.1]$





- we give the average peak height, and estimate the error by jackknife method.

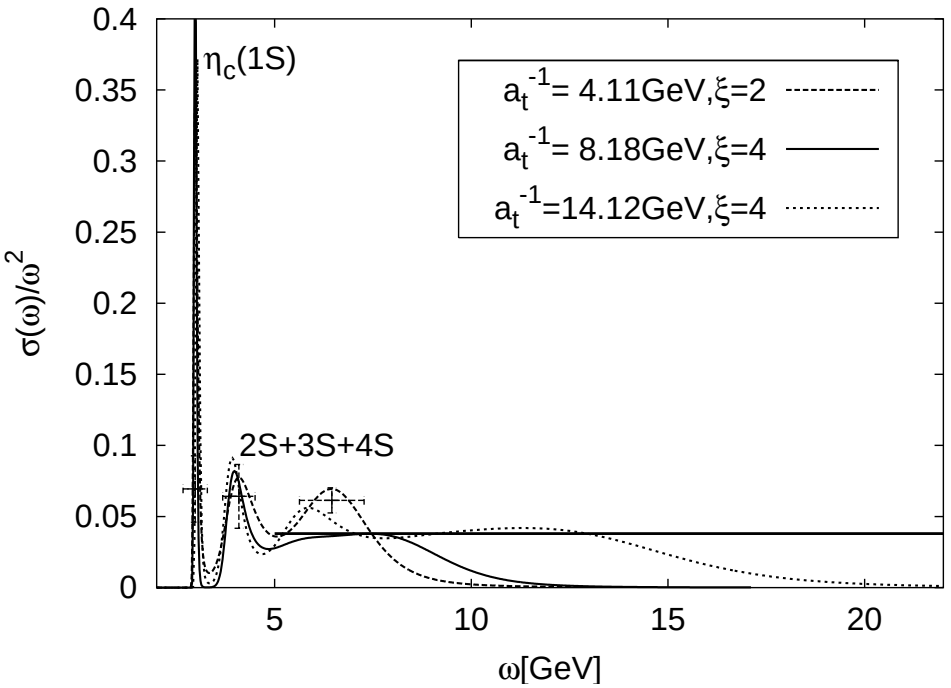
Outlines

Numerical simulation of QCD

(A.J., P. Petreczky, K. Petrov, A. Velytsky, Phys.Rev.D75:014506,2007)

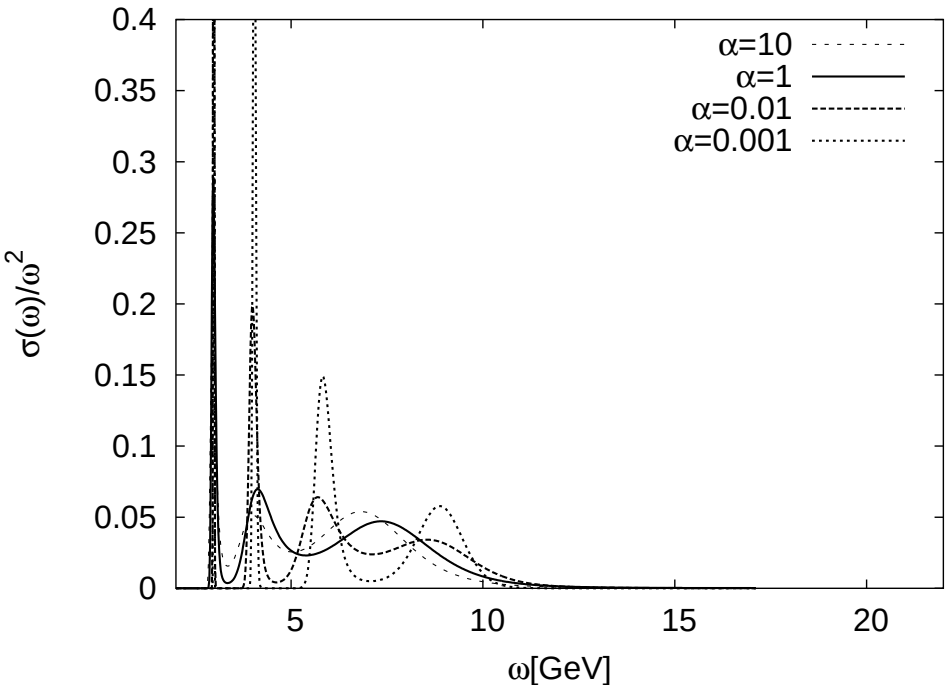
- Quenched QCD, gauge sector with standard Wilson action
- anisotropic lattice (Fermilab formulation), bare anisotropy parameter $\xi = 4$, tuned to reveal 4D rotational invariance
- lattice sizes: from $8^3 \times 64$ to $24^3 \times 32 \times 160$.
- measured mesonic operators in different quantum channels (S,PS,V,AV) for heavy quarks (c,b), different spatial wave functions (S and P-wave)
- for heavy quarks: anisotropic clover action, tree level clover coefficients
- for physical results see A. Velytskys talk

Zero temperature spectral function in the PS channel:



- horizontal error bars: peak regions
- vertical error bars: jackknife error
- horizontal line: free massless limit
- 1S state (η_c) is clean, second peak sum of excited states
  $\Rightarrow$ larger peak integral

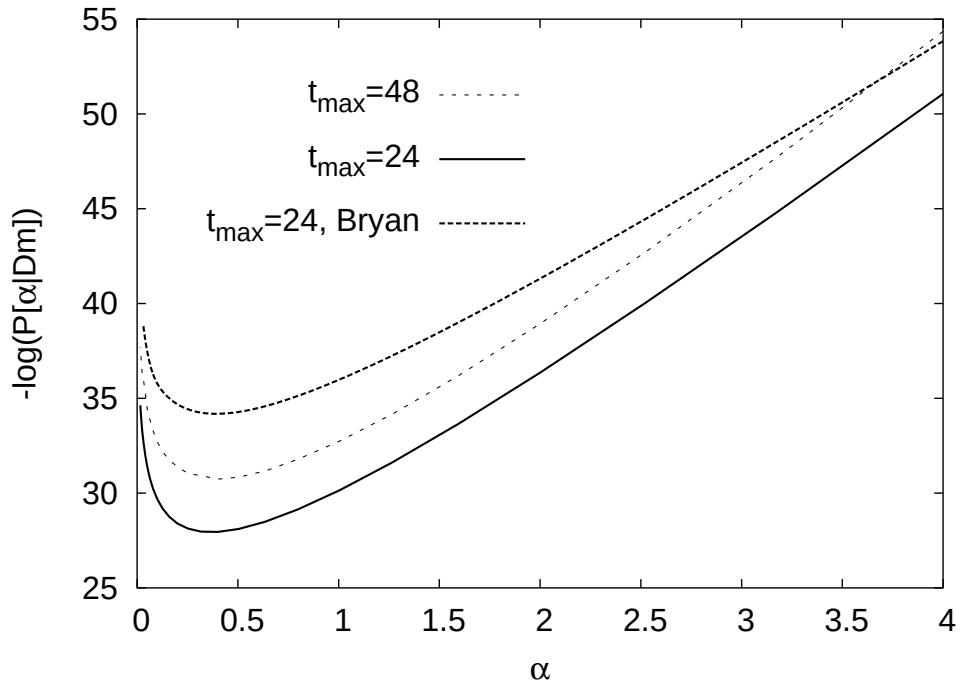
α-dependence of the result:



- same qualitative structure
- position of physical peaks are approximately α-independent

Comparing our algorithm and the Bryan algorithm:

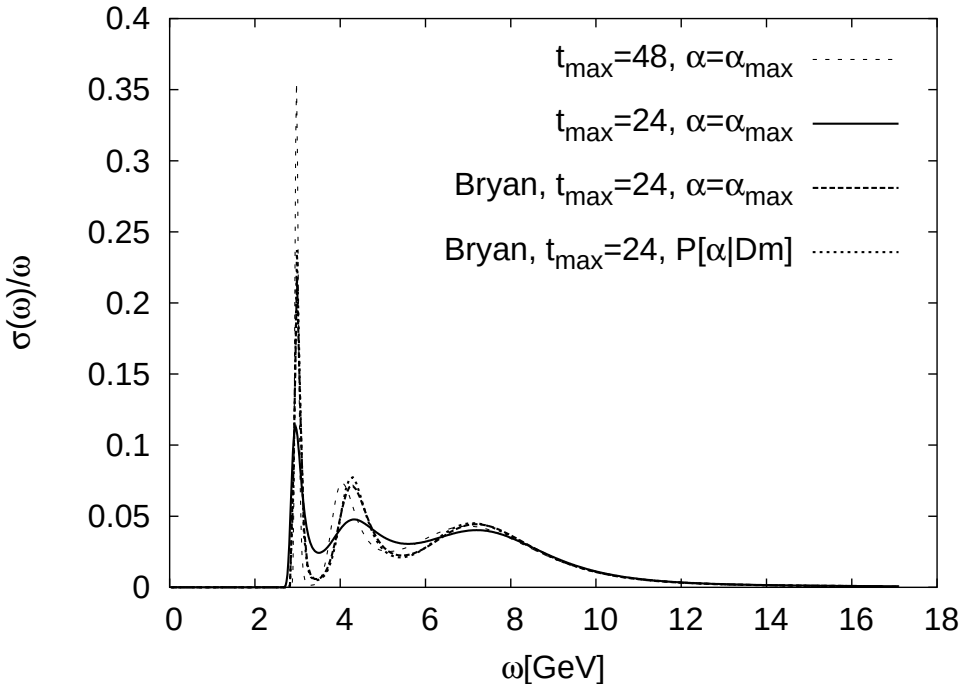
- numerically more stable, faster
- the α-probability function:



(almost) same curve with constant shift

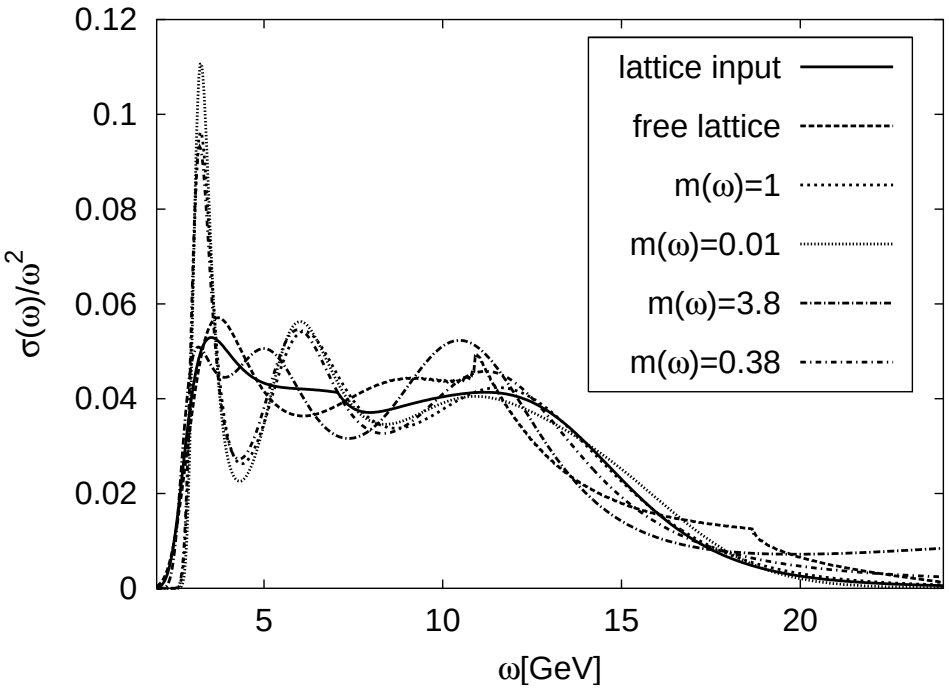
effect of the null-space of the SVD in the Bryan method

- spectral function:



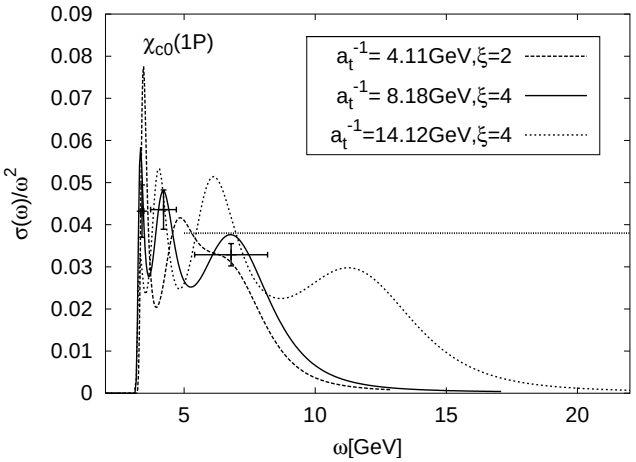
good agreement

Default model dependence:

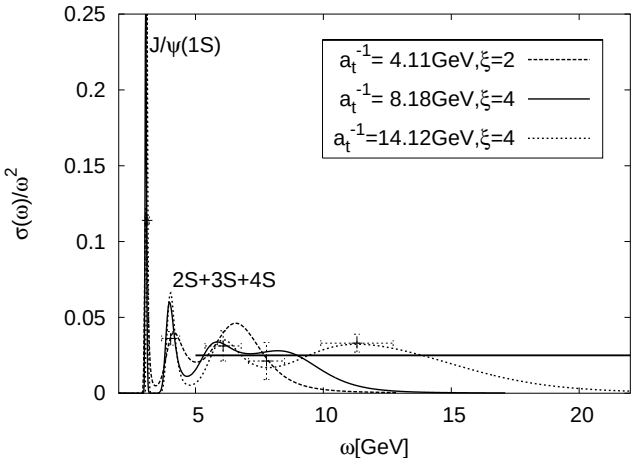


- it can be significant with widely different choices
- $m(\omega) = 1$ are best tested

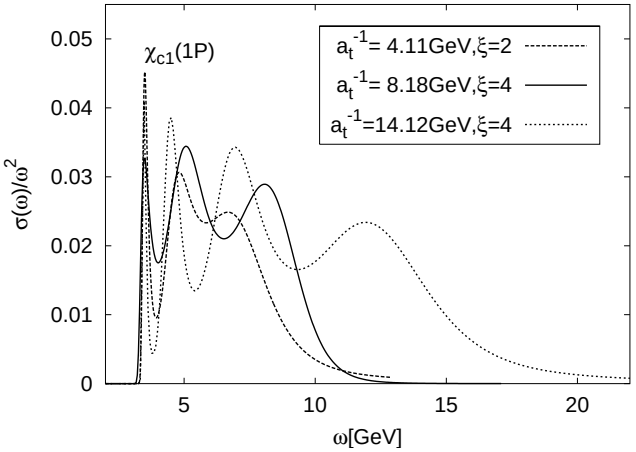
Other zero temperature spectral functions:



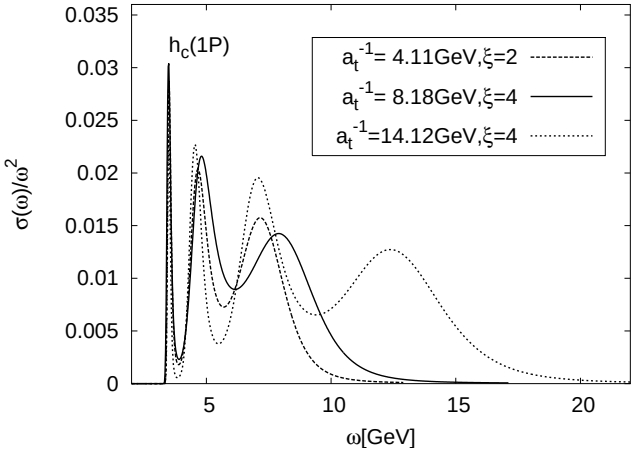
scalar



vector



axialvector



tensor

Conclusions:

- MEM is powerful method to obtain spectral functions from lattice QCD simulations
- can be rephrased as a potential minimization problem in a number-of-data dimensional space – faster, more reliable algorythm than Bryan algorithm
- peaky structures:
  - lower states are well reproduced
  - finite resolution
  - robust characterization: average peak value
- default model dependence requires more clarification